Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 154 (2019) 622–628

Procedia Computer Science

www.elsevier.com/locate/procedia

8th International Congress of Information and Communication Technology, ICICT 2019

Comparative Study of Various Feature Extraction Techniques for Pedestrian Detection

Hiranmai M[a], Niranjana Krupa B[a,]*, Nagaraj H K[b]

[a]PES UNIVERSITY, Bengaluru-85, India

[b]Robert Bosch Engineering and Business Solutions Private Limited, Bengaluru, India

Abstract

This paper presents, feature extraction techniques such as center symmetric local binary pattern (CSLBP), extended CSLBP (XCSLBP), speeded-up robust feature (SURF) with 64 and 128 feature descriptors and histograms of oriented gradients (HOG) applied on a set of images from INRIA person database, to detect pedestrians. About fifteen feature sets created using different combinations of the aforementioned methods are compared using two detectors, random forest (RF) and support vector machine (SVM). Performance validation is done based on the accuracy, precision, recall and space required for storing feature vectors. Experimental results have shown that CSLBP and the novel XCSLBP+CSLBP feature sets yield 100% accuracy, when used with RF classifier, whereas, the novel SURF-128+XCSLBP combination and SVM linear classifier gave 99.2% accuracy in detecting pedestrians.

Keywords: Pedestrian detection; center symmetric local binary pattern; speeded up robust feature; extended center symmetric local binary pattern; HOG

* Corresponding author. Tel.: +91-80-26724413
*E-mail address:*bnkrupa@pes.edu

1877-0509
10.1016/j.procs.2019.06.098

1. Introduction

Pedestrian detection is the most challenging intelligent system applications that would be helpful in advanced driver-assistance systems (ADAS). This need to detect pedestrians is quenched with various recently proposed feature extraction and classification techniques, which can be used in the separating pedestrian images from non-pedestrian images. In the past, researchers have proposed several approaches for pedestrian detection and object detection in general. Histograms of oriented gradients (HOG)[1] is one such feature extraction method, when applied in combination with linear support vector machine (SVM) on MIT pedestrian dataset close to perfect results were obtained. Furthermore, the performance of HOG was proved better using another dataset named 'INRIA', consisting of human images with different backgrounds and pose variations. The progress of pedestrian detection techniques in a span of ten years was studied in detail[2] where HOG was the most widely used technique in combination with other methods. Around 40 plus feature extraction methods were tested on several databases, including INRIA and Caltech USA, using three different categories of detectors; deformable part model (DPM), deep networks, and decision forests. There it was proved that decision forest detector yields better results in detecting pedestrians.

Center symmetric local binary pattern (CSLBP)[3], a variant of local binary pattern (LBP), was employed and it proved to be an efficient feature descriptor when compared with scale-invariant feature transform (SIFT), since it was tolerant to changes in illumination and robust to flat image areas. LBP and its variants such as CSLBP and local ternary pattern (LTP) were used along with SVM classifier to detect pedestrians and the results were compared with that of HOG and its variants. The outcome proved that LBP variants in combination with SVM perform better[4]. Oriented CSLBP features were also used in detecting pedestrians[5] from thermal images with the help of random forest (RF) classifier, resulting in a robust algorithm. Speeded-up robust feature (SURF)[6] descriptor, inspired from SIFT was implemented to detect objects. The method yielded better performance when it comes to variations in images, robustness and distinctiveness. When HOG features were extracted from the clustered SURF points and fed to SVM for pedestrian detection, it was observed that pedestrians' exact positions were tough to find using SURF points. Therefore, shift with importance sampling[7] (SIS) scheme was applied on SURF points to achieve an improved efficiency. However, in one of the recent studies[8], stop sign categories have been recognized where it was substantiated that SURF descriptors with SVM classifier produces best average accuracy even when images were rotated and scaled.

Experimental comparisons were made among SIFT, HOG, SURF and oriented chamfer distance (OCD) feature extraction methods using SVM as a classifier. It was demonstrated that the proposed OCD approach has a great ability of recognizing pedestrian contours[9]. A mixture of HOG and extended center-symmetric local binary pattern (XCSLBP) features with linear SVM classifier proved to outperform HOG-LBP mixture[10]. It was also noted that the false alarm rate was minimum and pedestrian detection was better with HOG-XCSLBP. Again, HOG and LBP feature extraction methods and the combination of the two were tried along with SVM, RF, and DPM, for detecting pedestrians during the day as well as night, on visible and far infrared (FIR) images[11]. It was concluded that HOG+LBP features produce the best result. Another hybrid model where XCSLBP is used with W^4 algorithm proved to be the best model in detecting moving objects, compared to W^4 algorithm alone[12]. The most commonly used classifiers in pedestrian detection are RF and SVM. RF classifier was first created by Tin Kam Ho[13]. In their work, a method was proposed for both training data and unseen data where a tree-based classifier was constructed, whose capacity can be expanded arbitrarily for improving the accuracy. The other classifier is SVM, proposed by Corinna and Vapnik[14], where the authors have extended the study on linearly non-separable classes. Basically, a solution for two-group classification problem has been presented in the paper along with a solution for finding an optimal hyperplane that is used to separate classes with maximum distance from the optimal margin.

This paper provides an insight into those recent techniques which can help in achieving the best accuracy and at the same time ensure reduced space complexities while detecting pedestrians. In this process, the images were acquired from 'INRIA' person database. Equal number of pedestrian and non-pedestrian images are considered. The extracted features are classified using two different classifiers, they are, RF and SVM. Under SVM, three kernels were adopted, linear, polynomial and radial basis functional (RBF). The feature extraction techniques implemented for the study are HOG, SURF in two variants i.e., SURF-64 and SURF-128, CSLBP and XCSLBP. It is interesting to see how the novel combination of SURF and LBP variants helped in recording the near perfect accuracy.

2. Proposed Methodology

The implementation and validation of the algorithm follows a methodological process to achieve valid results. The images used in the work are from INRIA Person Database[1]. Four hundred images were randomly chosen out of which 200 images contained pedestrians and the remaining 200 were non-pedestrian images. First step is preprocessing, where the images are converted to desired size and desired color space. Later, various feature extraction techniques are applied on the images and the resulting feature vectors are given as input to the classifiers in order to detect pedestrians.

2.1. Preprocessing Techniques

Resizing: This step is important to maintain uniform dimension of an image in order to exercise equal time in computation of feature vectors. Here, the images are set to a dimension of (300,600).

Grayscale Conversion: All images are converted to grayscale. This helps in classifying the subjects with better accuracy as the contrast and structure information is maintained.

2.2. Feature Extraction Techniques

This section provides a brief insight on the various techniques adopted to extract features from the chosen 400 images, before they are fed as input to the classifiers.

HOG: Features extracted from HOG[1] are useful in detecting objects that converts pixel-based representation into a gradient based representation. These are often used in linear classification techniques. Given an image to extract HOG features, is divided into small regions, which are connected as cells. Histograms are computed from the pixels contained in each of these cells. These are based on 'gradient directions' or 'edge orientations. As per the orientation of the gradient, every cell is discretized into angular bins ranging from 0 to 180 degrees. These discretized values are contributed from each pixel contained in the cell. Adjacent cells are grouped together to form blocks. Histograms obtained from each block are then normalized to form block histogram. Thus, set of these block histograms represent an image feature descriptor. Different cell sizes can be adopted for extracting features from images. Here, cell size of [32X32] was chosen to produce a feature vector of dimension [400X4896], where '4896' corresponds to the number of gradient points extracted from input image.

SURF: This technique is widely used in object detection, image registration, 3D reconstruction or classification. Here, approximate Gaussian second derivative mask is applied to an image at many scales. The execution of SURF[6] is performed with the help of very basic Hessian matrix approximation which reduces computation time. Use of Hessian matrix is mostly preferred over any other blob detectors due to its higher computation speed and accuracy, also to eliminate ill-localised structures that are irrelevant. In an image $\boldsymbol{I}$ consider any point $\mathbf{x}$, the pixel location, then, $\mathbf{x} = (x, y)$. When $\boldsymbol{I}(x, y)$ is the image intensity the second derivative of it defines the Hessian matrix $H(\mathbf{x}, \sigma)$, as in (1):

$$H(x, y, \sigma) = \begin{bmatrix} I_{xx(x,y,\sigma)} & I_{xy(x,y,\sigma)} \\ I_{xy(x,y,\sigma)} & I_{yy(x,y,\sigma)} \end{bmatrix} \tag{1}$$

Where, Gaussian function of standard deviation 'σ ', helps in computing the derivatives I_{xx}, I_{xy}, I_{yy}. In the presence of local maxima in the determinant matrix, the interest points are determined. Later, the region around each interest point is extracted and this describes a SURF feature vector. This is an n-dimensional vector representing an object which is based on some measurements like colour, numerical or both applied on image features. Two different variants were used in the work; SURF-64 and SURF-128. The first variant, SURF-64 resulted in 64 features per image, hence, for 400 images the feature vector dimension is [400X64]. Using the other variant, SURF-128, 128 features are extracted thus the final feature vector is of dimension [400X128].

CSLBP: It is a variant of LBP, but more advantageous than LBP as it produces 16 binary patterns unlike LBP, which produces 256 patterns. It is illumination invariant and one of the better texture detection-based techniques. Here, center symmetric pairs of pixels[3] (Fig. 1(a)) are compared to find compact binary patterns using (2) and (3).

$$CSLBP_{R,N,T}(\mathrm{x,y}) = \sum_{i=0}^{(N/2)-1}[g(\mathrm{k_i} - \mathrm{k}_{\mathrm{i}+(\mathrm{N}/2)})2^i] \tag{2}$$

$$g(x) = \begin{cases} 1, & x \geq T \\ 0, & x < T \end{cases} \tag{3}$$

Where, the pixel gray values are k_i and $k_{i+N/2}$, the number of pixels that are equally spaced is set to $N=8$ and the radius of the circle is made $R=2$. The gray level difference threshold T is set to 0.01. The feature vector is of dimension, [400X16], for 400 images, where, 16 represents the number of columns with values of varied intensities from 0 to 15.

XCSLBP: It is an extended version of CSLBP, a variant from LBP. It improves background modelling by comparing a pair of symmetrical pixels with the pixel at the center of the image, using (4). Thus, it overcomes the drawbacks of its aforementioned variants in handling threshold values and obtaining the binary pattern[10] or the final decimal number value.

$$XCSLBP_{P,R}(c) = \sum_{i=0}^{(P/2)-1}[s(n_1(i,c) + n_2(i,c))2^i] \tag{4}$$

The threshold function is defined as in (5):

$$s(z_1 + z_2) = \begin{cases} 1 & z_1 + z_2 \geq 0 \\ 0 & otherwise \end{cases} \tag{5}$$

$n_1(i,c)$ and $n_2(i,c)$ are defined using (6) and (7), respectively.

$$n_1(i,c) = n_c + n_i - n_{i+(P/2)} \tag{6}$$

$$n_2(i,c) = (n_i - n_c)(n_{i+(P/2)} - n_c) \tag{7}$$

Where P represents the count of pixels that are equally spaced and R, the radius on which these pixels appear, and (xc, yc) is the center and in ni , i = 0,…,P-1. Both XCSLBP and CSLBP generates the histograms which will be of same length but XCSLBP avoids the selection of threshold values. This technique is also illumination invariant. Fig. 1(b) illustrates the computation method of XCSLBP.

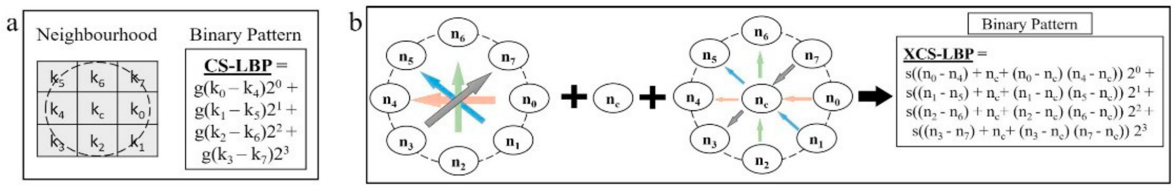

Fig. 1: (a) Shows the binary pattern computation for CSLBP method; (b) Shows the XCSLBP encoding method

2.3. Classification:

Two classifiers were adopted for classification of the images into pedestrian and non-pedestrian images, RF classifier and SVM with linear, polynomial and RBF kernels.

RF Classifier: A supervised learning algorithm based on the concept of creating a forest with decision trees. The class membership is calculated by repeatedly partitioning a dataset into uniform subsets. In a tree structure, every node represents a test on the feature, a branch is an outcome and a leaf node is the class. Higher the number of trees in forest, higher will be the accuracy. This algorithm is widely used because it handles missing values of the data and does not overfit the model in presence of more trees. It can also handle larger datasets with higher dimensionality[13]. Different number of decision trees are tested and is set to *100* as it resulted in the maximum accuracy.

SVM Classifier: By convention, SVM[14] is a discriminative classifier formally defined by building hyperplanes in a high dimensional space. When these hyperplanes have the largest margin a good separation of the data, into different categories, can be achieved. This algorithm provides the largest flexibility in partitioning separable ("linear") and non-separable ("non-linear") classes. These non-linear classes make use of kernel functions to classify the data in higher dimensional space. Here, three kernels are employed under this classifier, linear, polynomial of degree two and RBF kernel with c=1 and gamma=0.7. Performance accuracy depends on the hyperplane selection and kernel parameters.

3. Experimental Outcome and Discussion

By convention, validation is conducted by dividing the dataset into training and testing sets. Therefore, out of 400 images, 70% are used as training set, that constitutes 280 images, and the remaining 30% are used as testing set, comprising of 120 images. This technique also helps in fitting the best model with lowest error on all parts of the data. All the images are resized and converted to gray scale (Fig. 2(a)). On these training and testing sets, five main feature extraction techniques, HOG, SURF-64, SURF-128, CSLBP and XCSLBP are applied to obtain five sets of features. Later, these features sets are combined to form ten more feature sets, thus bringing the total number of feature sets to 15. The outcome of the five aforementioned feature descriptors is shown in Fig. 2(b) to Fig. 2(f). For every image input HOG resulted in 4896 features, SURF-64 and SURF-128 gave 64 and 128 features, respectively. Whereas, CSLBP and XCSLBP output had 16 features each. Each one of 15 feature sets is given as input to the classifiers, RF and SVM.

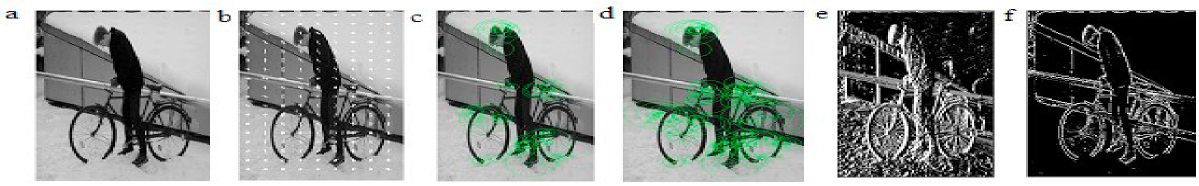

Fig. 2: (a) Gray image; (b) HOG descriptors; (c) SURF- 64 descriptors; (d) SURF-128 descriptors; (e) CSLBP features; (f) XCSLBP features.

Results obtained from the classifier models after testing are tabulated in Table. 1, where it can be seen that overall accuracies (AC), precision (PR) and recall (RE) percentages are provided for all the 15 feature sets. Although, RBF kernel in SVM classifier was employed to classify the images, the results procured were in the range of 50~55%, for all the 15 sets, even after varying the kernel parameters. Therefore, the results are not presented. For the other classifier models, RF, SVM-linear and SVM-polynomial, the accuracies obtained are in the range of 76.7~100%. RF classifier results show that CSLBP and XCSLBP+CSLBP have helped in getting 100% accuracy. It is also important to note that both the sets require less space to store features as they generate only 16 features each. The only drawback of RF is that it requires more training time and more data when compared to SVM. On the other hand, SVM-linear has given 99.2% accuracy and 99% PR and RE for both HOG+XCSLBP and SURF128+XCSLBP feature sets. Since, HOG+XCSLBP feature vector length is 4912 when compared to that of SURF128+XCSLBP, which is only 144, it can be highlighted that the latter approach is better. With SVM-polynomial classifier of degree two, the feature set SURF64+XCSLBP has given the highest accuracy of 96.6% and 97% of PR and RE. In this case the length of the feature vector is only 80. On the basis of the results obtained, combinations of SURF variants and XCSLBP are better suited for pedestrian detection with SVM classifiers. Whereas, with RF classifier, CSLBP and XCSLBP+CSLBP are the best sets of features.

4. Conclusion

In this paper, features were extracted from 400 gray images using HOG, CSLBP, XCSLBP, SURF-64 and SURF-128 techniques. The resulting five feature sets were combined to create ten more sets. These feature sets were fed as input to the RF, SVM with linear, polynomial and radial basis function kernels. The results of all the classifier models for all the fifteen input feature sets were compared, taking into account the accuracy, precision, recall and the memory space. Based on these parameters, it can be concluded that CSLBP and CSLBP+XCSLBP feature sets when used as input to RF classifier yield 100 % results. In addition, if time taken by RF for training is considered, SVM-linear with SURF128+XCSLBP is a better approach and it also helped in gaining near-perfect accuracy of 99.2%. As future work, the proposed pedestrian detection system will be tested on more number of real-time samples captured under different lighting conditions which would help in assisting the drivers in any ADAS.

Table 1. Results obtained from classifiers in (%) for different combinations of feature descriptors

Feature Extraction Technique	Feature Vector Dimension	SVM – Linear			SVM – Polynomial			RF		
		AC	PR	RE	AC	PR	RE	AC	PR	RE
HOG	[400 * 4896]	97.5	98	97	90.8	91	91	86.6	87	87
CSLBP	[400 * 16]	93.3	94	93	90	90	90	100	100	100
SURF-64	[400 * 64]	86.6	87	87	79.2	82	79	76.7	77	77
SURF-128	[400 * 128]	92.5	93	93	80.8	84	81	81.7	82	82
XCSLBP	[400 * 16]	98.3	98	98	93.3	94	93	97.5	98	97
HOG+CSLBP	[400 * 4912]	94.2	95	94	90	90	90	96.6	97	97
HOG+SURF-64	[400 * 4960]	97.5	98	97	79.2	82	79	93.3	93	93
HOG+SURF-128	[400 * 5024]	97.5	98	97	80.8	84	81	95	95	95
CSLBP+SURF-64	[400 * 80]	80.8	83	81	87.5	88	88	99.2	99	99
CSLBP+SURF-128	[400 * 144]	83.3	85	83	87.5	88	88	93.3	94	93
SURF-64+SURF-128	[400 * 192]	95	95	95	85	86	85	80.8	81	81
HOG+XCSLBP	[400 * 4912]	99.2	99	99	94.2	94	94	90.8	91	91
SURF-64+XCSLBP	[400 * 80]	95.8	96	96	96.6	97	97	90.8	91	91
SURF-128+XCSLBP	[400 * 144]	99.2	99	99	92.5	93	93	92.5	93	93
XCSLBP+CSLBP	[400 * 32]	84.2	85	84	90.8	91	91	100	100	100

5. References

1. Navneet D, Bill T. Histogram of oriented gradients for human detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2005; p. 886 - 93.
2. Rodrigo B, Mohamed O, Jan H, Bernt S. Ten years of pedestrian detection, what have we learned?. European Conference on Computer Vision (ECCV), 2014; p. 613–27.
3. Heikkilä M, Pietikäinen M, Schmid C. Description of interest regions with center-symmetric local binary patterns. Proceedings International Conference on Computer Vision, Graphics and Image Processing, 2006; p. 58-69.
4. Yongbin Z, Chunhua S, Richard H, Xinsheng H. Effective pedestrian detection using center-symmetric local binary/trinary patterns. Computer Vision and Pattern Recognition, arXiv:1009.0892v2, September 2010.
5. Mi Ra Jeong, Jun-Yong Kwak, Jung Eun Son, Byoung Chul Ko, Jae-Yeal Nam. Fast pedestrian detection using a night vision system for safety driving. 11th International Conference on Computer Graphics, Imaging and Visualiztion, 2014; p. 69-72.
6. Herbert B, Andreas E, Tinne T, Luc Van Gool. SURF: Speeded up robust features. Computer Vision and Image Understanding (CVIU), 2008; 110(3):346-59.
7. Bo-Yuan Wong, Jun-Wei Hsieh, Li-Chih Chen, Duan-Yu Chen, Hong-Yi Liu, Chong-Po Liao. Pedestrian detection using SURF and Shift with importance sampling. IEEE International Conference on Consumer Electronics, Taiwan, 2014; p. 201-02.
8. Loussaief S, Abdelkrim A. machine learning framework for image classification. 7th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT), 2016; p. 58-61.
9. Dandi Chen, Siyu Xia, Yuan Zhou. Pedestrian detection via contour fragments. Proceedings of the 35th Chinese Control Conference, July, 2016; 4054-60.

10. Vasuki P, Veluchamy S. Pedestrian detection for driver assistance systems. Fifth International Conference on Recent Trends in Information Technology, 2016.
11. González A, Fang Z, Socarras Y, Serrat J, Vázquez D, Xu J, López AM. Pedestrian detection at day/night time with visible and FIR cameras: A comparison. *Sensors*, 2016; 16(6):820.
12. Hatim Derrouz, Azeddine El Hassouny, Rachid Oulad Haj Thami, Hamid Tairi. Hybrid method for background modeling and subtracting. Intelligent Systems and Computer Vision (ISCV), 2017; p. 1-5.
13. Tin Kam Ho. Random decision forests. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, QC, 1995; p. 278–82.
14. Corinna Cortes, Vladimir Vapnik. Support vector networks. *Machine learning*.1995; 20:273-97.